Reprinted from: *Karger Gazette* No. 13, 2 (March 5, 1966)

ISI Eases Scientists' Information Problems Provides Convenient Orderly Access to Literature

On the fifth floor of a modern office building, a stone's throw from Philadelphia's Independence Hall, some 125 associates of the Institute for Scientific Information are busy providing services and turning out publications specifically designed to help scientists cope with their literature and information problems. In addition to providing these services, now used by more than 80,000 scientists, ISI devotes a large percentage of its budget to research and development of new techniques for handling scientific information.

ISI's pioneering research has nurtured one of the most promising new concepts for information retrieval developed during the past 50 years – the concept of *citation indexing*.

The *Science Citation Index*, a comprehensive, multi-disciplinary Index to science, is based on this concept, and is published quarterly and annually by ISI. The *Science Citation Index* enables librarians, administrators and scientists to retrieve both scientific and technological information in a manner heretofore not thought possible. Although the *SCI* system has been in current operation for only two years, it is gaining wide acceptance by academic, industrial and government researchers.

For 1965 alone, approximately 300,000 source items were indexed to generate the *SCI* for that calendar year. This involved the processing of some 3,000,000 cited references originally written by approximately one-third of a million authors.

The basic concept of citation indexing is based on the observation that when one article cites another article there must be a subject relationship between the two articles. Each bibliographic citation as, e.g., 'A.R. JONES, Fed. Proc. *53*, 1493 (1959)', is an unambiguous descriptor or symbol which designates the subject matter discussed in some aspect of the citing work. For this reason, citation indexing provides the scientist a unique and efficient method for conducting, in a minimum of time, highly specific searches, since each cited reference becomes an entry point in the Index.

The *Science Citation Index* enables one to trace the literature forward in time; that is, to go from an earlier cited article to a later citing article. This contrasts with conventional systems in which one usually locates a current article and builds up a bibliography by tracing backward in time through the footnotes cited in the article at hand. *SCI* indexing is based on the author's rather than the indexer's estimate of the subject content of an article. For this reason, the *SCI* has been said to be particularly responsive to the user's, rather than the indexer's viewpoint. For the time being *Science Citation Indexes* are available only for the years 1961, 1964 and 1965. The 1966 Index is, of course, now in production. ISI plans to publish Indexes for 1962 and 1963 as soon as practical.

During the past eight years, ISI has initiated such services as *Current Contents of Chemical, Pharmaco-Medical and Life Sciences* and *Current Contents of Space, Electronic and Physical Sciences*, which are estimated to be read by over 60,000 scientists each week. In addition, ISI introduced the *Index Chemicus* in 1960, a unique abstracting-indexing service that pinpoints and reports graphically on approximately 120,000 newly synthesized chemicals each year. The *Index Chemicus* is estimated to be read regularly by 10,000 synthetic chemists, pharmacologists and others throughout the world – at least 50% outside the United States.

In 1965, ISI introduced the *Automatic Subject Citation Alert*. *ASCA* is, in part, based on the citation indexing principle and is the first large scale computer-based service commercially available for selective dissemination of information to individual scientists and/or groups that covers all major disciplines.

One of the major precepts of ISI's philosophy is that the various fields of science and technology are, in fact, interdisciplinary and should not be categorized into arbitrary a priori classifications. Many of the world's literature abstracting organizations continue to operate on an a priori basis, but ISI believes that it is not proper to try to rigidly define the perimeters of fast-changing areas of research. Highly interdisciplinary fields such as molecular biology, oceanography, behavioral science, etc., require that information be extracted simultaneously from the literature of all disciplines.

ISI is attempting to *broaden* the *base* of coverage of all its services while at the same time making it easier for the individual scientist to *narrow* down and extract that *facet* of multi-disciplinary information of specific interest and value to him.

It has been estimated from various sources that some 50,000 scientific and technical journals are being published in more than 65 languages. Estimates from ISI indicate that less than 1,000 journals account for over 90% of the significant scientific reports – approximately 250,000 articles per year – of which a considerable portion is duplication in various forms. This means that, contrary to popular belief, a relatively small number of journals account for a high percentage of the significant articles published. ISI's present journal coverage already exceeds 1,500 journals.

The offices of ISI are on the fifth floor of this modern building in Philadelphia.

While the processing of several hundred thousand items each year is no trivial problem, ISI's studies have led to both manual and computer systems that enable the scientist to overcome the sense of helplessness that is conveyed by unqualified allusions to 50,000 journals – many of which ceased publication long ago, or publish a trivial number of significant papers. While there has been a steady growth of the literature, ISI believes the so-called "information explosion" can indeed be controlled. Contrary to the general cliché that "there is too much scientific information", ISI would assert that "there is often little or no information available about a specific problem". This "needle in the haystack problem" is faced almost daily by every scientist. In many such cases, conventional literature searching techniques may not be practical. However, such techniques as the *Science Citation Index* and *ASCA* enable the searcher to find out quickly if there is – or is not – information available on a specific point. By reducing the time required to do comprehensive but specific searches, ISI hopes to increase the use of the literature as an effective research instrument.

Information science as practiced by conventional libraries and information centers has not usually been thought to be an important or exciting part of research and development by scientists. However, with the increased opportunities for productive exploitation of the scientific literature, this situation is now changing. Computer technology has also had great impact on the field, making large scale information handling systems practical. More than half of ISI's total working force consists of information scientists, librarians and data processing technicians. An IBM 360 and an IBM 7044 system are used to process approximately 100,000 new punch cards per week. The cards are then converted to magnetic tape. Using special programs developed at ISI, these tapes are scanned automatically for scientists all over the world.

ISI's all-consuming goal is to organize the world's total output of significant scientific and technical literature into an integrated file and provide convenient current and retrospective access to the file by individual scientists. By the end of 1966, ISI will have processed approximately ten million reference records. The goal of total retrospective coverage for the literature of the twentieth century should be completed in gradual stages over the next decade. This enormous file will be of inestimable value both for current laboratory research and also for new types of historical studies.

While the file building is going on, ISI's research program, headed by Dr. Irving H. Sher, who received his doctorate in biochemistry from Johns Hopkins University, is experimenting with new techniques on the commercial services already being provided by ISI. He is working hand in hand with A.W. Elias, ISI's Director of Publications and Services. The fruits of this research already have started to pay off in the new *SCI Searching Service* in which it is possible to obtain a retrospective search of the literature on a specific topic using ISI's data bank. Searches are performed daily at low cost–often by telephone or cable request.

Even at the present time, ISI is supplying copies of its magnetic tapes to various organizations for their internal use. These tapes are used in many ways including citation and word-searching approaches to subject matter retrieval. A new approach which simplifies the complexities of existing word-profile schemes is now in its final stages of testing and debugging. This additional approach will augment the *ASCA* citation profiles and enable scientists to construct word profiles to search literature for concepts that do not lend themselves as readily to the citation approach. The combined approaches through words, authors, organizations, citations and journals will provide a depth of indexing to satisfy very exacting requirements at low cost.

Studies on the sociological implications of science and scientific literature are an important part of ISI's R & D program. A recent report for the US Government on *The Use of Citation Data in Writing the History of Science* [9] is looked upon by many in the field as having unique application to pinpointing areas for future scientific research, identifying the research fronts of science, and evaluating published work by providing convenient access to criticisms by peers.

Supporting all of ISI's retrieval and dissemination activities is an original system of storing scientific journals. ISI learned from the outset that it must back up any bibliographic listing in *Current Contents*, *Index Chemicus*, etc., with convenient access to the papers listed. It is generally expected that the scientist will use local library facilities for this purpose. In addition, reprints are frequently requested directly from authors by using the computer-produced address directories which ISI publishes. Nevertheless, there is frequently the need for the scientist to obtain quickly a paper that is not received at his institution. For this reason, ISI operates the *Original Article Tear Sheet* service. In the *OATS* service, tear sheets are literally torn from extra copies of journals received by ISI. Although many requests are received at ISI, only a small number of requests are for any given article. This enables the Institute to satisfy most needs from a maximum of five or six copies of each issue.

ISI takes great pride in the good relations it has maintained with hundreds of publishers throughout the world who cooperate to minimize the many inherent time lags in publication and recognize that the growth of interdisciplinary science requires that individuals have access to many journals they do not ordinarily scan directly. The growth in popularity of ISI's services has been paralleled by a comparable increase in journal circulation throughout the world.

The key to effective retrieval of information is to convert the process of information discovery to that of information recovery. By effectively disseminating information in the first place, as through *Current Contents*, *Index Chemicus*, and *ASCA*, the memory of the individual scientist can be better exploited to recover that information when it is needed again. On the other hand, since the individual scientist can only be exposed to a fraction of the world's information in his lifetime, he also needs what has been described as "systematic serendipity" – an organized process of information discovery of that which he did not know existed. Finally, he must be able to quickly and efficiently determine if it is valid to conclude that *no* information is available and he can proceed in the laboratory with a reasonable degree of assurance that unwitting duplication of efforts will be avoided.

The *SCI* and *ASCA* serve these two latter functions. The above objectives of information science indicate the important creative and service roles the entire information processing profession can play in future research and development activity throughout the world.

References

1. Garfield, E.: Science citation index. A new dimension in indexing. Science *144:* 649–654 (1964).
2. Martyn, J.: An examination of citation indexes. Aslib Proceedings *17:* 184–196 (1965).
3. Garfield, E.: Science citation index. Answers to frequently asked questions. Revue Internationale de la Documentation *32:* 112–116 (1965).
4. Sher, I. H. and Garfield, E.: New tools for improving and evaluating the effectiveness of research. Paper presented at the Second Conference on Research Program Effectiveness, Washington, D.C., July 27–29, 1965 (Office of Naval Research). To be published in Proceedings of the Conference.
5. Price, D. J. de: Networks of scientific papers. Science *149:* 510–515 (1965).
6. Berndt, H.: Das Institut für wissenschaftliche Information in Philadelphia. Dokumentation *9:* 174–177 (1962).
7. Garfield, E. und Stevens, L. J.: Über den Science-Citation-Index (SCI) und verwandte Entwicklungen der jüngsten Zeit. Nachrichten für Dokumentation *16:* 130–139 (1965).
8. Garfield, E.: Citation Indexing: A natural science literature retrieval system for the social sciences. American Behavioral Scientist *7:* 58–61 (1964).
9. Garfield, E.; Sher, I.H. and Torpie, R. J.: The use of citation data in writing the history of science (Philadelphia: Institute for Scientific Information 1964), p. 75.

INSTITUTE FOR SCIENTIFIC INFORMATION
325 Chestnut Street Philadelphia Pa 19106 USA

6-00-65-05